

Hepatocellular Carcinoma Prediction Using Large Language Model And Meditron

Suleiman Zeila Dubow

Department Of Computing And Information Science Kenyatta University

ABSTRACT

This project investigated the integration of machine learning, large language models (LLMs), and the Meditron medical language model to enhance early prediction and interpretation of hepatocellular carcinoma (HCC) in resource-constrained healthcare settings such as Kenya, where late diagnosis and limited access to advanced diagnostic tools remain major challenges. Using a design science and experimental research approach, an AI-driven clinical decision-support prototype was developed to process unstructured clinical conversations, extract relevant medical features, predict HCC risk, and generate clinically interpretable explanations. A Random Forest classifier was trained on a balanced secondary clinical dataset comprising equal proportions of HCC and non-HCC cases, with preprocessing techniques including median imputation and feature scaling applied to ensure data quality and model robustness. The trained model achieved an accuracy of 0.8780, an F1-score of 0.8780, and an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.9452, demonstrating strong predictive performance. We applied an LLM to convert free-text clinical notes into structured features suitable for the trained predictive model allowing automated HCC risk prediction from conversational input. Then the prediction outputs comprising probabilistic risk scores were submitted to the Meditron model that gave medically grounded interpretations of the results by contextualizing anticipated risk in respect to known clinical indications of HCC. We integrated LLM-based data extraction, machine learning-based prediction, and Meditron-driven interpretation, which enhanced transparency, explainability, and clinical relevance of the system. Results demonstrate that hybrid AI architectures that combine predictive modeling and domain-specific reasoning with medical language can provide accurate, interpretable, and scalable tools for early HCC risk assessment with great potential to support clinical decision-making and improve cancer care delivery in low-resource healthcare settings.

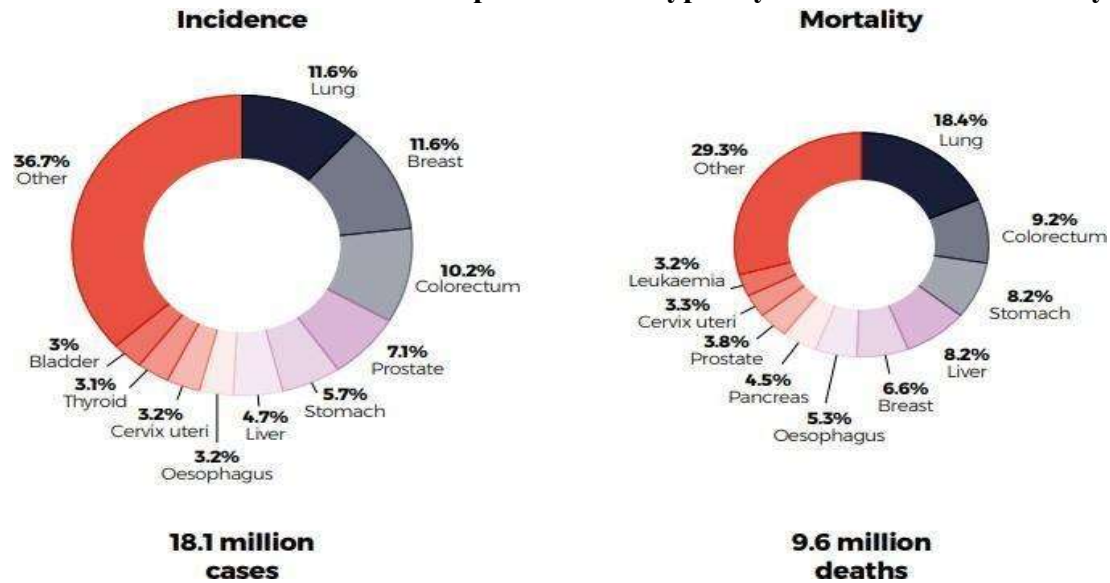
CHAPTER ONE

INTRODUCTION

1.1 Background of the study

Cancer is a complex, heterogeneous class of multivariable diseases involving the uncontrollable multiplication of malignant cells and tumor spread (Pravin & Sudhir, 2018). Since cancer is the major cause of premature death in 134 out of 183 countries worldwide, it is a major public health concern (World Health Organization, 2020). According to the World Health Organization (2020), there were approximately 9.6 million cancer-related deaths and 18.1 million new cases of cancer worldwide in 2018. As seen in the figure below, the ten most common cancers account for 60-70% of cancer incidence and death (World Health Organization, 2020).

Figure 1.1: Global Distribution of the Top 10 Cancer Types by Incidence and Mortality (2018)



Note: The top ten cancer types globally in 2018, based on incidence and death rates for both sexes, are provided in World Health Organization’s 2020 WHO REPORT ON CANCER (World Health Organization, 2020).

Some of the principal carcinogens operating on humans are tobacco, asbestos, aflatoxins and UV radiation. There is increasing evidence of lifestyle variables that lead to cancer, for example food, physical exercise and alcoholic intake (World Health Organization, 2020). Today’s treatment options can be surgery, systemic therapies (like immunotherapy, endocrine therapy and chemotherapy) and radiotherapy, alone or in combination, depending on the disease stage, patient’s health status and tumor characteristics, for the best outcomes. (Stewart et al., 2003; World Health Organization, 2020)

The incidence of primary liver cancer, primarily hepatocellular carcinoma, appears to be increasing in Western countries (Befeler & Di Bisceglie, 2002; Loy et al., 2022). Primary liver carcinomas (Loy et al., 2022) comprise 75-85% of hepatocellular carcinoma. Hepatocellular carcinoma has been described as a disease with a high but underappreciated burden in sub-Saharan Africa, with age-standardized incidence rates as high as 41 per 100,000 per year, male predominance, an earlier age of onset than in Western populations and a very poor prognosis in that most patients die within a year of symptom onset. This pattern is mostly related to chronic hepatitis B and aflatoxin exposure (Kew, 2013). Kenya is therefore part of this high incidence East African region where hepatocellular carcinoma is one of the leading causes of cancer related death and has the same epidemiological profile of early onset, male predominance and strong association with hepatitis B infection despite the lack of exact national statistics due to limited data from cancer registries (Kew, 2013). In some parts of Africa, Asia and South America, a high incidence of hepatocellular carcinoma is still driven by aflatoxin exposure and chronic infection with hepatitis B viruses (HBV) (Stewart et al., 2003).

The high incidence of hepatocellular carcinoma (HCC) in sub-Saharan Africa, more so in Kenya, is not complemented with early identification which remains a severe clinical and systemic concern. Delayed diagnosis is affected by limited availability to current imaging modalities like CT and MRI, inadequate surveillance of high-risk populations (e.g., those with chronic hepatitis B), and lack of skilled healthcare workers (Asafo-Agyei et al., 2023; Kew, 2013).

Other barriers to early diagnosis include the absence of standardised screening programs, poor health infrastructure and low awareness among the general population, leading to late presentation in most patients and thus limiting the efficacy of contemporary available therapies (Stewart et al., 2003; World Health Organization, 2020). These limitations highlight an urgent need for affordable, scalable early prediction strategies to improve HCC detection and patient outcomes in resource-limited settings like Kenya.

1.2 Statement problem

Early identification and prediction of hepatocellular carcinoma (HCC) remains a major concern especially in low- and middle-income countries such as Kenya where the burden of the illness is high but with little healthcare resources. Current diagnostic methods such as blood biomarkers include alpha-fetoprotein (AFP) and imaging modalities such as CT and MRI have great limitations. AFP has low specificity and variable sensitivity and imaging modalities are expensive and generally unavailable in resource poor countries resulting in delayed diagnosis and poor patient outcomes (Befeler & Di Bisceglie, 2002; Loy et al., 2022). Furthermore, the risk factors for HCC, including chronic hepatitis B infection, aflatoxin exposure, alcohol use and metabolic problems, are complicated, multivariate and closely intertwined. This complexity makes standard statistical and clinical methods for early prediction worthless. Machine learning (ML) models have shown potential in classifying cancer based on high-dimensional clinical, genomic and epigenomic data but they tend to lack interpretability and transparency limiting their acceptability and trust in clinical decision making (Kamita et al., 2024).

Large language models (LLMs) have demonstrated amazing ability in processing medical knowledge, understanding natural language and clinical reasoning. However, they are not widely used in oncology as they are mostly built for text-based tasks rather than structured predictive modeling and lack relevant knowledge in the cancer domain (M. Li et al., 2024; Nazi & Peng, 2024). This represents a critical gap in developing an integrated, explainable and cost-effective system that combines the predictive powers of machine learning and the contextual reasoning capabilities of LLMs to enable early HCC detection, particularly in resource-constrained settings such as Kenya.

1.3 Objectives

To develop and evaluate an ensembled artificial intelligence framework integrating machine learning and medical large language models for hepatocellular carcinoma risk prediction and clinical interpretation.

1.3.1. The specific objectives

- To examine the limitations of existing diagnostic and predictive approaches for hepatocellular carcinoma, particularly in resource-constrained healthcare settings.
- To develop an ensemble model combining a Large Language Model (LLM) and the Meditron Medical Large Language Model for clinical feature extraction and medical reasoning.
- To integrate random forest with the developed ensemble to form another ensemble model.
- To evaluate the performance of the final ensemble.

1.4 Justification

The rising cancer burden in Kenya, particularly hepatocellular carcinoma (HCC), highlights the pressing need for improved diagnostic and therapeutic strategies. The rising incidence and high mortality rates linked to late-stage diagnosis create an urgent need for novel approaches that can improve early diagnosis and eventually patient outcomes (World Health Organization, 2020; Kamita et al., 2024).

Furthermore, current diagnostic approaches, such as advanced imaging and biomarker detection, are often expensive and not accessible to many, particularly in resource-constrained settings. This highlights the need for cost-effective, scalable, and rapid diagnostic approaches that can be implemented broadly without the need for specialized facilities or expertise (Befeler & Di Bisceglie, 2002; Loy et al., 2022).

These challenges could be mitigated by AI, which facilitates early detection through data-driven predictive models, reduces diagnostic costs, and helps physicians in decision-making. The application of AI technologies, especially machine learning and large language models, has the potential to revolutionize healthcare by improving diagnostic accuracy, access, and timely care for HCC (M. Li et al., 2024; Nazi & Peng, 2024).

1.5 Scope of the study

The study was conducted to develop and evaluate an artificial intelligence-based approach for early detection of hepatocellular carcinoma (HCC), with a particular focus on resource-constrained settings like Kenya. The study used secondary data sources, including clinical, demographic, and risk factor information related to HCC, to build and assess predictive skills. The focus was on integrating machine learning techniques with a large language model (Meditron) to enhance both the accuracy of predictions and the contextual clinical reasoning.

The study was limited to the analysis of structured and semi-structured data on identified risk factors of HCC (hepatitis B infection, aflatoxin exposure, patient history) and did not include real-time clinical application and prospective testing of patients. The model was also designed as an aid in early detection and not to replace the standard diagnostic methods such as imaging or laboratory tests, but to supplement the clinical judgment.

Specifically, the study focuses on the following areas:

- **Geographical Focus** – The study is focused on healthcare systems and cancer care practices in Kenya.
- **Technological Integration** – Development and integration of a custom LLM with the Meditron platform for predictive modeling and disease analysis.
- **Patient Data Utilization** – Analysis of patient records and data to enhance symptom interpretation and diagnostic accuracy.
- **Evaluation Metrics** - Assessment of the proposed solution in terms of diagnostic sensitivity, accuracy and specificity and comparison with existing approaches to establish its effectiveness.
- **Target Audience** – The study is relevant for healthcare professionals, policy makers and researchers interested in improving cancer diagnostics in low resource settings.

The study also aimed to assess model performance, interpretability, and feasibility in a low-resource setting, rather than attempting to integrate the model into a full hospital system or implement policy. Limitations such as data availability, generalizability across populations, and computational resources were also considered, defining the scope to which the findings and conclusions of this study are relevant.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction to cancer

Cancer is a broad category of diseases defined by the uncontrolled growth and spread of abnormal cells in the body. Cell division and death are closely regulated processes in normal physiological conditions; however, cancer disrupts these regulatory mechanisms, resulting in the creation of malignant tumors that can invade surrounding tissues and spread to distant organs (Pravin & Sudhir, 2018). Genetic

abnormalities, epigenetic changes, and environmental exposures all contribute to unregulated growth, which disrupts normal cellular function.

Cancer is a serious public health concern around the world, accounting for one of the main causes of illness and mortality. According to the World Health Organization (2020), cancer caused around 10 million deaths and nearly 19 million new cases worldwide, with a large proportion happening in low- and middle-income nations. Cancer is becoming more prevalent as a result of population growth, aging, and greater exposure to risk factors including as tobacco use, poor diets, physical inactivity, infections, and environmental toxins (World Health Organization, 2020).

Cancer is a diverse illness with over 100 distinct varieties, each with their own biological characteristics, risk factors, and clinical consequences. Surgery, chemotherapy, radiation, immunotherapy, and targeted treatments are all common treatment options, which are frequently used in combination depending on the kind and stage of the disease (Stewart et al., 2003). Despite breakthroughs in diagnosis and treatment, early detection is still crucial for improving survival rates since many malignancies can be treated more effectively when detected early on.

2.2 Hepatocellular carcinoma (liver cancer)

Hepatocellular carcinoma (HCC) is the most frequent kind of primary liver cancer, accounting for 75-85% of all liver cancers globally (Loy et al., 2022). It is most usually associated with chronic liver disease and cirrhosis, which are caused by persistent infections with hepatitis B virus (HBV) or hepatitis C virus (HCV), as well as exposure to environmental carcinogens such as aflatoxins (Kew, 2013; Stewart et al., 2003). HCC etiology is a multistep process characterized by persistent inflammation, hepatic injury, regeneration, and the accumulation of genetic and epigenetic changes that eventually lead to malignant transformation.

Clinically, HCC is frequently asymptomatic in the early stages, with symptoms such as stomach pain, weight loss, jaundice, and liver dysfunction showing until later in the disease. Ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) are common imaging modalities used in diagnostics, as are serum biomarkers such as alpha-fetoprotein. However, these methods are limited in sensitivity and accessibility, especially in low-resource settings (Befeler & Di Bisceglie, 2002). As a result, many patients are detected at advanced stages, when curative treatments like surgical resection, liver transplantation, or local ablation are no longer feasible, emphasizing the essential need of early identification and improved predictive tools.

2.3 Large language models (LLMs)

Large language models (LLMs) are powerful artificial intelligence systems trained on massive volumes of textual material to understand, create, and reason in natural language. LLMs in the medical arena have proven great capabilities in tasks such as clinical recordkeeping, information extraction, decision support, and question answering, making them increasingly useful in cancer research and healthcare delivery (M. Li et al., 2024; Nazi & Peng, 2024). Their ability to synthesize enormous amounts of biomedical knowledge and patient-related data makes them ideal tools for assisting doctors in complex decision-making processes.

In oncology, LLMs are being investigated for applications such as cancer risk assessment, early detection support, treatment suggestion, and patient education. They have combined several data sources, including as clinical notes, academic literature, and patient-reported symptoms, to find trends that would be difficult to spot using traditional methods. This is especially crucial in diseases like hepatocellular carcinoma (HCC), where early-stage symptoms are frequently vague or non-specific, and identification is strongly

dependent on the combination of many risk factors and clinical signs (Q. Li et al., 2024). LLMs have also help comprehend complex genomic and epigenomic data, resulting in improved cancer categorization and staging (Kamita et al., 2024).

Despite their potential, the use of LLMs in cancer, including HCC, is hampered by various obstacles. Many existing models have not been extensively trained on specialized oncology datasets, resulting in domain-specific knowledge gaps and poor clinical dependability (M. Li et al., 2024). Furthermore, LLMs are primarily intended for natural language processing activities and may lack robust predictive skills when used alone for illness diagnosis. Concerns about explainability, data protection, and clinical validation also impede broad implementation in healthcare settings (Nazi & Peng, 2024). As a result, there is an increasing demand for domain-adapted LLMs, like as Meditron, that may be combined with predictive models to improve early diagnosis and clinical decision support for diseases like hepatocellular carcinoma.

2.4 pre-trained models on cancer

Pre-trained models have become a significant advancement in artificial intelligence, particularly in medical domains where labelled datasets are frequently scarce. These models are trained on large-scale general or domain-specific datasets and then fine-tuned for specific downstream tasks like classification, prediction, and information extraction. Pre-trained models have been frequently used in cancer research to increase diagnostic accuracy, survival prediction, and disease classification using transfer learning from huge biological corpora (Kamita et al., 2024).

Pre-trained models are increasingly being utilized to examine complicated and high-dimensional data in hepatocellular carcinoma (HCC) and other liver malignancies, such as genomes, epigenomics, and clinical records. Studies like Classification of Early and Late-Stage Liver Hepatocellular Carcinoma Patients from Their Genomics and Epigenomics Profiles show that pre-trained machine learning architectures may effectively distinguish between disease stages by detecting hidden patterns in molecular data. These models improve prediction performance when compared to classic statistical methods, especially in early-stage detection, where clinical signs are frequently vague or non-specific.

The following table shows Pre-trained and AI Models Used in Cancer Research:

Table 2.1: Pre-trained and AI Models Used in Cancer Research

Model	Model Type	Cancer-Related Application	Reported Use in Research
BioBERT	Biomedical Large Language Model (LLM)	Biomedical text mining, cancer literature analysis, clinical entity extraction	Used for cancer-related natural language processing tasks including oncology information extraction from biomedical literature and clinical records.
ClinicalBERT	Clinical Transformer Model	Cancer diagnosis support, clinical note analysis, EHR interpretation	Applied in oncology clinical note classification and extraction of cancer-related patient information from electronic health records.
PubMedBERT	Biomedical Transformer Model	Cancer text classification and biomedical knowledge extraction	Trained on PubMed articles and widely used in cancer research NLP applications.

GPT-4	Large Language Model	Oncology reasoning, cancer report summarization, medical question answering	Used experimentally for oncology clinical decision support and cancer-related medical reasoning tasks.
Meditron	Medical Large Language Model	Clinical reasoning, explainable healthcare AI, oncology support	Applied in medical dialogue generation and explainable AI systems for healthcare interpretation.
ResNet	Deep Convolutional Neural Network	Tumor detection, histopathology analysis, cancer imaging classification	Commonly used in breast cancer, lung cancer, liver cancer, and skin cancer imaging studies.
DenseNet	Deep CNN Architecture	Cancer image classification and tumor segmentation	Used in radiology and histopathological cancer image analysis.
InceptionV3	Deep CNN Architecture	Breast cancer detection and cancer stage classification	Applied in medical imaging-based cancer diagnosis systems.
VGG16	Deep CNN Architecture	Histopathological cancer classification	Used for cancer cell classification and microscopy image analysis.
U-Net	Deep Learning Segmentation Model	Tumor segmentation and medical image processing	Extensively used in cancer lesion segmentation and radiotherapy planning.
XGBoost	Ensemble Machine Learning Model	Cancer survival prediction and recurrence analysis	Used in structured clinical data prediction for liver, breast, and lung cancers.
Random Forest	Ensemble Decision Tree Model	Cancer risk prediction and prognosis modeling	Applied in hepatocellular carcinoma, breast cancer, and colorectal cancer prediction studies.
DeepSurv	Deep Neural Survival Model	Cancer survival analysis and prognosis prediction	Used for personalized cancer survival prediction using clinical variables.
PathAI	AI Pathology Model	Cancer pathology analysis and tumor identification	Applied in digital pathology and cancer biopsy interpretation.
Watson for Oncology	Oncology Clinical Decision Support System	Cancer treatment recommendation and oncology support	Used in oncology decision support and cancer treatment planning.

However, despite their usefulness, pre-trained models in cancer have some drawbacks. Many models are task-specific and necessitate substantial fine-tuning to accommodate different cancer kinds or populations, which can be difficult in low-resource environments. Furthermore, difficulties such as lack of interpretability, data bias, and limited generalization across varied groups restrict their therapeutic utility

(Kamita et al., 2024). In the context of HCC, these limitations underscore the need for more adaptable and explainable AI systems that can combine clinical knowledge with data-driven predictions to aid in early diagnosis.

2.5 Random Forest in Healthcare

Random Forest has been widely used in cancer prediction tasks such as breast cancer classification, lung cancer detection, prostate cancer risk analysis, and hepatocellular carcinoma (HCC) stage prediction due to its ability to handle high-dimensional, nonlinear, and heterogeneous clinical datasets (Classification of Early and Late-Stage Liver Hepatocellular Carcinoma Patients from Genomics and Epigenomics Profiles, 2023). It is particularly effective in medical domains because it reduces overfitting through ensemble learning and performs well with noisy, missing, or imbalanced data, which are common in real-world cancer datasets (Kamita et al., 2024). In addition, Random Forest provides feature importance scores that help identify key risk factors such as hepatitis infection, liver enzyme levels, and demographic variables, making it more interpretable compared to many black-box models. Therefore, it is selected in this study as the core machine learning model for HCC prediction because it offers a strong balance between accuracy, robustness, and interpretability, which is essential for clinically reliable and explainable cancer diagnosis systems.

2.6 Meditron models

Meditron models are a collection of open-source big language models created exclusively for the medical domain by conducting large-scale continuous pre-training on curated biomedical and clinical corpora. According to MEDITRON-70B: Scaling Medical Pretraining for Large Language Models and MEDITRON: Open Medical Foundation Models Adapted for Clinical Practice, these models are created by taking strong general-purpose foundation models (such as LLaMA-based architectures) and training them on large medical datasets such as PubMed articles, clinical guidelines, and de-identified clinical notes. Compared to conventional LLMs, this domain-adaptive pretraining allows the models to develop deeper medical reasoning capabilities, higher terminology knowledge, and better alignment with clinical workflows (MEDITRON-70B, 2024; MEDITRON, 2024).

Meditron's training technique involves continuous pretraining on large-scale biomedical corpora, followed by instruction tailoring to increase performance on medical question answering, clinical reasoning, and decision-support tasks. The Meditron-70B variant, in particular, illustrates that increasing model size combined with domain-specific pretraining improves performance on biological benchmarks. Meditron outperforms general LLMs without medical adaptation, such as ChatGPT or LLaMA, in terms of clinical query accuracy, contextual comprehension of symptoms, and reduced irrelevant or nonmedical responses. However, it still uses static training data and does not include real-time patient-specific structured prediction unless integrated with external models (MEDITRON-70B, 2024).

Meditron outperforms general-purpose LLMs in medical reasoning and clinical knowledge tasks thanks to its specialized training corpus, but it may still fall behind task-specific machine learning models in structured prediction tasks like disease classification or risk scoring. Traditional machine learning models trained on structured datasets (e.g., genomic or clinical tabular data) frequently achieve higher predictive accuracy for specific outcomes such as HCC stage classification, whereas Meditron excels at interpretability and contextual explanation of medical concepts (M. Li et al., 2024). Meditron outperforms other medical LLMs like Clinical Camel and BioGPT in clinical reasoning and instruction-following because of its greater scale and more complete pretraining pipeline (Nazi & Peng, 2024; MEDITRON, 2024).

Overall, Meditron serves as a bridge between general-purpose language models and domain-specific clinical intelligence, making it especially useful when combined with predictive machine learning systems for complex diseases like hepatocellular carcinoma, which require both structured prediction and clinical reasoning.

2.7 Benefits of Meditron

Meditron models are useful in healthcare because they have a strong domain-specific understanding acquired through large-scale biomedical and clinical pretraining, allowing for more accurate interpretation of medical terminology and clinical context than general-purpose LLMs (MEDITRON-70B, 2024; MEDITRON, 2024). They improve clinical decision support by combining patient data with scientific knowledge to help with reasoning and diagnosis, which is especially useful in complicated diseases like hepatocellular carcinoma, where early symptoms are frequently non-specific (Nazi & Peng, 2024). Meditron also enhances access to medical expertise in resource-limited situations and, when paired with machine learning models, promotes explainability by offering human-readable clinical interpretations of prediction outputs (M. Li et al., 2024).

2.8 Integrated LLMS and Meditron

The combination of large language models (LLMs) with Meditron results in a hybrid framework that combines machine learning's predictive power with superior medical reasoning capabilities. In this configuration, LLMs may process and structure patient-related information, while Meditron provides domain-specific biomedical knowledge to increase output interpretation and clinical relevance (M. Li et al., 2024; MEDITRON, 2024). This integration is particularly effective in hepatocellular carcinoma (HCC), where complicated risk variables and non-specific symptoms require both precise prediction and appropriate clinical explanation. By integrating prediction with explanation, the integrated method improves diagnostic support, interpretability, and clinical usability, particularly in low-resource healthcare settings (Nazi & Peng, 2024; MEDITRON-70B, 2024).

2.9 Knowledge Gap

Reference	Study Focus	Key Finding	Identified Knowledge Gap
Befeler & Di Bisceglie (2002)	Role of AFP in HCC diagnosis	AFP has variable sensitivity (76%–91%) and very low specificity (9%–32%)	Biomarker-based screening is unreliable for early HCC detection
Loy et al. (2022)	Diagnosis and epidemiology of HCC	Imaging (CT/MRI) improves diagnosis but is mainly effective in later stages	Early-stage HCC remains difficult to detect using imaging alone, especially in low-resource settings
Kew (2013)	HCC in sub-Saharan Africa	HCC is highly prevalent with late presentation and poor survival outcomes	Lack of early detection systems and limited screening infrastructure in Africa
Kamita et al. (2024)	AI in cancer prediction	Machine learning improves cancer prediction using clinical and genomic data	Existing models struggle with generalization and real-world deployment in LMICs

Reference	Study Focus	Key Finding	Identified Knowledge Gap
M. Li et al. (2024)	Large language models in medicine	LLMs perform well in medical reasoning and NLP tasks	LLMs lack strong structured predictive capability for clinical diagnosis
Nazi & Peng (2024)	ChatGPT and medical LLM use	LLMs show promise in clinical decision support but require careful regulation	Limited trust, explainability, and validation in high-risk clinical environments
Classification of Early and Late-Stage Liver HCC Patients from Genomics and Epigenomics Profiles (2023)	Genomic-based HCC classification	ML can distinguish early vs late HCC using molecular data	High computational complexity and limited clinical interpretability
MEDITRON-70B (2024)	Scaling medical LLMs	Domain-adapted LLMs improve medical knowledge understanding	Meditron lacks direct predictive modeling capability for disease outcomes
MEDITRON: Open Medical Foundation Models Adapted for Clinical Practice (2024)	Clinical adaptation of LLMs	Meditron improves clinical reasoning and biomedical QA	Still requires integration with structured models for accurate prediction tasks

Existing literature demonstrates that biomarkers (AFP), imaging approaches, machine learning models, and big language models do not provide a complete solution for early hepatocellular carcinoma identification. Biomarkers are inaccurate, imaging is costly and requires late-stage intervention, ML models lack interpretability and deployment resilience, and LLMs lack structured prediction power. Even Meditron, which excels at medical reasoning, cannot reliably forecast sickness on its own. This is an obvious need for an integrated, explainable AI system that combines prediction and clinical reasoning for early HCC identification in resource-constrained environments like Kenya.

CHAPTER THREE RESEARCH METHODOLOGY

3.1 Research design

This study used Design Science Research (DSR) methodology and an experimental research design to create and test an ensemble-based artificial intelligence model for predicting hepatocellular carcinoma (HCC). The artifact created is a hybrid ensemble model that combines Random Forest (as the primary predictive component) and a Meditron-based large language model (LLM) to improve interpretability and clinical decision support. The DSR approach was appropriate because it focuses on creating and testing an IT artifact that addresses a real-world healthcare issue, specifically early HCC identification in resource-constrained contexts.

The design science method was iterative, with steps that included problem identification, artifact design,

system construction, ensemble model integration with the Meditron LLM, and structured evaluation. The Random Forest component performed the core prediction task by learning from structured clinical and risk-factor data, whereas Meditron interpreted outputs and provided natural language explanations to aid clinician understanding. This integration results in an ensemble decision-support system that incorporates predictive learning and medical reasoning.

The experimental design component was used to evaluate the performance of the developed ensemble system using HCC-related datasets. The model was trained, tested, and validated using standard evaluation metrics including accuracy, sensitivity, and specificity. Comparative experiments were conducted against standalone machine learning models and traditional diagnostic approaches to assess performance improvements achieved through the ensemble integration. Since the Meditron has already been developed, the focus of this study is on evaluation and validation of the ensemble model for early HCC prediction and clinical decision support in Kenya and similar low-resource settings.

3.2 Target Population

This study's target population comprised of hepatocellular carcinoma (HCC)-related clinical datasets that included demographic information, laboratory findings, medical history, symptoms, and liver cancer risk factors for prediction and clinical explanation. The study concentrated on patient-related healthcare data pertinent to HCC diagnosis, which included both structured clinical characteristics and unstructured free-text medical descriptions.

Hepatocellular carcinoma is the most frequent type of primary liver cancer, accounting for 75-85% of all cases globally (Villanueva, 2019). In 2020, liver cancer killed more than 830,000 people worldwide, making it one of the major causes of cancer-related death (Sung et al., 2021). In Sub-Saharan Africa, the burden of HCC remains significantly high due to the prevalence of hepatitis B virus (HBV) and hepatitis C virus (HCV) infections, limited early screening, and delayed diagnosis (Yang et al., 2019).

The study population included clinical records containing major HCC-related variables such as HBV infection, HCV infection, alcohol use, liver cirrhosis, fatigue, abdominal pain, weight loss, alpha-fetoprotein (AFP) levels, liver enzyme indicators, and demographic information such as age and gender. These variables were chosen because prior research has established them as significant indicators of HCC risk and development (Akinyemiju et al., 2017).

The target population also includes free-text clinical narratives and patient symptom descriptions, which were used in the Large Language Model (LLM) feature extraction method. These textual inputs enabled the study to assess the LLM's capacity to extract medically important information, convert it into structured JSON representations, and facilitate machine learning prediction and Meditron-based clinical explanation production.

As a result, the selected cohort supplied clinically relevant and diversified healthcare data, which was required for creating and evaluating the proposed hybrid HCC prediction system.

3.3 Sampling and Sampling Procedure

This study used purposive sampling to select clinical datasets linked to hepatocellular carcinoma (HCC) that were relevant to the research aims. Purposive sampling was deemed appropriate since the study particularly sought patient records that included medically important variables linked with liver cancer diagnosis and prediction.

The sampling approach centered on selecting datasets comprising structured and unstructured clinical information related to HCC risk factors and progression. Clinical records containing variables such as hepatitis B virus (HBV), hepatitis C virus (HCV), alcohol consumption, liver cirrhosis, fatigue, abdominal

pain, liver enzyme levels, alpha-fetoprotein (AFP), age, and gender were included because they have been linked to the development of hepatocellular carcinoma (Villanueva, 2019).

To improve dataset quality and predictive reliability, records with inadequate clinical information, large numbers of missing values, or irrelevant features were eliminated. This preprocessing procedure guaranteed that the machine learning and Large Language Model (LLM) components were trained on clinically relevant data appropriate for HCC prediction and explanation creation.

The dataset was partitioned into training and testing subsets to allow for experimental validation of the created model. Stratified sampling approaches were used during dataset splitting to maintain the proportional distribution of HCC-positive and HCC-negative instances. This method minimized sampling bias and increased the Random Forest classifier's generalization capabilities during prediction.

Purposive sampling was also performed on free-text clinical records for the LLM feature extraction component. These records were chosen to assess the system's capacity to detect medically relevant elements and transform them into structured JSON and dataframe formats for predictive analysis.

As a result, the selection approach ensured that the selected datasets were clinically relevant, representative, and appropriate for assessing the performance of the proposed hybrid HCC prediction framework combining Random Forest and Meditron algorithms.

Purposive sampling was also performed on free-text clinical records for the LLM feature extraction component. These records were chosen to assess the system's capacity to detect medically relevant elements and transform them into structured JSON and dataframe formats for predictive analysis.

As a result, the selection approach ensured that the selected datasets were clinically relevant, representative, and appropriate for assessing the performance of the proposed hybrid HCC prediction framework combining Random Forest and Meditron algorithms.

3.4 Data Collection

Secondary hepatocellular carcinoma (HCC)-related clinical datasets from publicly accessible healthcare repositories on Kaggle were used in this investigation. The chosen datasets included structured clinical and demographic factors related to the diagnosis and prognosis of liver cancer. Because secondary data offered easily accessible, clinically relevant records acceptable for machine learning experiments, model training, and validation, it was a reasonable choice for this investigation.

Age, gender, hepatitis B virus (HBV) and hepatitis C virus (HCV) status, alcohol consumption history, liver cirrhosis indicators, liver enzyme measurements, alpha-fetoprotein (AFP) levels, tumor-related information, and other clinically significant HCC risk factors were all included in the gathered dataset. These factors were chosen because prior research has shown them to be important indicators of the onset and spread of hepatocellular carcinoma (Villanueva, 2019).

The Large Language Model (LLM)-based feature extraction component of the study was supported by free-text clinical descriptions and symptom narratives in addition to structured data. The algorithm was able to replicate real-world clinical encounters, where patient data is frequently entered in an unstructured manner, thanks to these textual inputs.

For preprocessing, transformation, model construction, and assessment, the gathered data was kept in the experimental setting.

3.5 Preparing Data

The obtained HCC dataset had to be cleaned up, organized, and prepared for machine learning in order to be used for clinical explanation generation and predictive modeling. In order to enhance dataset quality,

minimize inconsistencies, and guarantee compliance with the suggested hybrid AI framework, this step was required.

The data preparation process consisted of several stages:

I. Data Cleaning

Incomplete, duplicated, and inconsistent records were identified and handled during preprocessing. Missing numerical values were imputed using mean imputation expressed as:

$$x_i = \frac{1}{n} \sum_{j=1}^n x_j$$

where:

x_i represents the imputed value

X_j represents observed feature values

n represents the total number of non-missing observations

Records containing excessive missing or irrelevant information were excluded from the dataset to improve predictive reliability.

II. Outlier Detection and Handling

To find unusual clinical observations that can have a detrimental impact on machine learning performance, outlier detection was used. Alpha-fetoprotein (AFP) levels and liver enzyme readings are examples of numerical variables where outliers were found using the Interquartile Range (IQR) approach.

The IQR was computed as follows:

$$IQR = Q_3 - Q_1$$

where:

Q_1 represents the first quartile

Q_3 represents the third quartile

Outliers were identified using the conditions:

$$x < Q_1 - 1.5(IQR) \quad \text{or} \quad x > Q_3 + 1.5(IQR)$$

Detected outliers were either capped or removed depending on their clinical relevance and impact on model stability.

III. Data Transformation and Encoding

Machine-readable numerical representations were created from clinical variables that had categorical values, such as gender, alcohol intake, and hepatitis status. This change made it possible for machine learning algorithms to efficiently handle categorical healthcare data. Additionally, the JSON format of the structured clinical variables that the Large Language Model (LLM) collected from free-text input was converted into structured dataframe representations appropriate for machine learning research.

IV. Data Normalization

In order to decrease scale variability among features and enhance predictive accuracy, numerical variables such as liver function tests and alpha-fetoprotein (AFP) values were standardized using Min-Max normalization.

The normalization process was expressed as:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

where:

x represents the original feature value

x_{\min} represents the minimum feature value

x_{\max} represents the maximum feature value

x' represents the normalized value

V. Feature Selection

To increase computational efficiency and predictive accuracy, redundant and low-importance characteristics were eliminated while clinically significant variables linked to HCC prediction were included.

3.6 Modelling

This phase involved developing, training, evaluating, and integrating machine learning and Large Language Model (LLM) components for predicting hepatocellular carcinoma (HCC) and providing clinical explanations. To develop an understandable HCC prediction system, the proposed framework combined traditional machine learning techniques with transformer-based medical language understanding.

The architecture starts with the ingestion of the HCC dataset, followed by several preprocessing stages, including:

- Missing value handling.
- Outlier detection
- Feature Encoding
- Feature Scaling

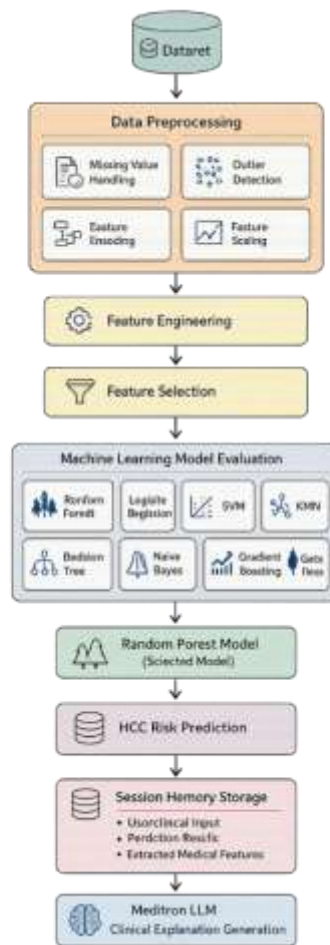
Following preprocessing, feature engineering and feature selection were used to create clinically relevant predictive variables for machine learning classification. The transformed dataset was then fed into different supervised machine learning algorithms for comparison.

The following machine learning models were tested: Random Forest, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Naïve Bayes, Gradient Boosting, and Extra Trees Classifier. Comparative experimental evaluation revealed that the Random Forest classifier outperformed all other models. Consequently, Random Forest was chosen as the final predictive model for HCC risk classification.

Following prediction, the resulting HCC risk score, extracted medical data, and user clinical input were saved in a session-based memory structure. The Meditron Large Language Model then collected the recorded session data and created clinically appropriate natural language explanations to aid healthcare decision-making and interpretation.

The architecture therefore integrates:

- Data preprocessing
- Machine Learning Prediction
- Session-based context preservation.
- LLM-based medical explanation generating.



Three subsets of the dataset were created in order to carry out the predictive modeling process:

- 70% of the Training Set
- 20% Validation Set
- 10% of the Test Set

The machine learning models were fitted and trained using labelled HCC clinical data from the training dataset. The algorithms discovered trends and connections between clinical characteristics and HCC outcomes at this phase.

Model performance was optimized and hyperparameters were tuned using the validation dataset. To increase predictive accuracy and decrease overfitting, parameters like the number of estimators, maximum tree depth, regularization parameters, and learning configurations were changed.

The train-validation-test splitting strategy follows established machine learning best practices for reducing overfitting and underfitting while improving model reliability and generalization performance in healthcare prediction systems.

The Random Forest classifier was set up with:

- 300 estimators
- Maximum depth of 12
- Balanced class weighting
- Random-state initialization of 42

The implemented models were defined as follows:

Model Name	Algorithm Type	Key Parameters	Purpose in Study
Random Forest	Ensemble Learning	n_estimators=300 max_depth=12 class_weight="balanced" random_state=42	Used as the final selected model for HCC prediction due to superior predictive performance, robustness, and ability to handle nonlinear relationships and imbalanced clinical data.
Logistic Regression	Linear Classification	max_iter=1000	Used as a baseline probabilistic classification model for comparison against ensemble approaches.
Support Vector Machine (SVM)	Margin-Based Classification	kernel="rbf" C=1 probability=True	Applied to model nonlinear decision boundaries within HCC clinical data using the Radial Basis Function kernel.
K-Nearest Neighbors (KNN)	Instance-Based Learning	n_neighbors=7	Used to classify HCC cases based on similarity between patient clinical records.
Decision Tree	Tree-Based Learning	max_depth=8	Used to generate interpretable rule-based classification structures for HCC prediction.
Naïve Bayes	Probabilistic Classification	Default Parameters	Applied as a probabilistic learning approach based on conditional independence assumptions between features.
Gradient Boosting	Ensemble Boosting	n_estimators=200	Used to improve predictive performance through sequential error correction during model training.
Extra Trees	Ensemble Learning	n_estimators=300 random_state=42	Evaluated as an alternative ensemble approach using randomized tree construction for improved variance reduction.

Random Forest prediction was mathematically represented as:

$$\hat{y} = \text{mode} \{T_1(x), T_2(x), \dots, T_n(x)\}$$

where:

- $T_i(x)$ represents predictions from individual decision trees
- \hat{y} represents the final predicted HCC class

During model training, bootstrap aggregation and randomized feature selection were applied to improve generalization and reduce overfitting. The splitting criterion within decision trees utilized Gini Impurity expressed as:

$$Gini = 1 - \sum_{i=1}^n p_i^2$$

where:

P_i represents the probability of a class within a node

The study used the trained models to generate predictions on the testing dataset following training and validation. Comparative analysis showed that Random Forest outperformed the other assessed algorithms in terms of prediction. The model maintained good prediction accuracy and generalization capabilities while managing class imbalance, heterogeneous healthcare data, and nonlinear clinical correlations.

Following prediction, a session-based memory structure was used to store the computed HCC risk results, extracted clinical data, and user input. After retrieving the stored session data, the Meditron Large Language Model produced natural language explanations for healthcare decision support that were clinically significant.

3.7 Assessment and Information Analysis

Evaluation and data analysis included figuring out the best method for predicting hepatocellular carcinoma and evaluating the predictive performance of the created machine learning models.

To maintain the class distribution between HCC-positive and HCC-negative cases, the processed dataset was split into training and testing subsets using stratified sampling techniques. The classifiers Random Forest, Logistic Regression, SVM, KNN, Decision Tree, Naïve Bayes, Gradient Boosting, and Extra Trees were then compared.

The following assessment metrics were used in the performance analysis:

- **Accuracy**
- **Precision**
- **Recall (Sensitivity)**
- **F1-score**
- **Specificity**
- **Area Under the Curve (AUC-ROC)**

Prediction accuracy was calculated using:

- **Accuracy:** The proportion of correctly predicted cases.

$$Accuracy = (True\ Positives + True\ Negatives) / Total\ Samples$$

- **Precision:** The ratio of real positive forecasts to all positive predictions.

$$Precision = True\ Positives / (True\ Positives + False\ Positives)$$

- **Recall (Sensitivity):** The ratio of real positive forecasts to all actual positive cases

$$Recall = True\ Positives / (True\ Positives + False\ Negatives)$$

- **F1-score:** The harmonic-mean of precision and recall.

$$F1\text{-score} = 2 * (Precision * Recall) / (Precision + Recall)$$

- **AUC-ROC:** Area under the Receiver Operating Characteristic curve, which assesses the model's overall performance across multiple thresholds.

Experimental analysis demonstrated that Random Forest achieved superior predictive performance compared to the other evaluated models. The model effectively handled nonlinear clinical relationships, heterogeneous healthcare data, and class imbalance while maintaining strong generalization capability. The final evaluation also assessed the effectiveness of the Meditron-generated clinical explanations in improving interpretability and supporting clinical decision-making within the proposed hybrid HCC prediction framework.

3.7 Ethical consideration

In this study, we strictly respected and developed ethical principles, research rules, and data protection standards that were maintained throughout the study due to the sensitive nature of healthcare information.

Secondary datasets from publicly accessible healthcare libraries and research platforms were used in this study. In order to eliminate personally identifying information (PII), such as patient names, identification numbers, contact information, and other identifiers that can jeopardize patient privacy and confidentiality, all patient-related data included in the study was anonymized.

To maintain data confidentiality and integrity, all datasets and experimental results were stored in secure and password-protected environments, with access restricted to only approved research workers. Data handling processes were created to reduce the risks associated with illegal access, disclosure, or misuse of healthcare information.

The Kenyatta University Ethics Review Committee approved the study before it began. In addition, the study was granted an official research license by the National Commission for Science, Technology, and Innovation (NACOSTI), in accordance with Kenyan legislation governing scientific research involving healthcare and human-related data.

The study also promoted responsible usage of artificial intelligence technologies such as the Random Forest model and the Meditron Large Language Model (LLM). The generated forecasts and clinical explanations were intended solely for research and decision-support purposes, not to substitute expert medical diagnosis or clinical judgment.

By following ethical approval requirements, NACOSTI research regulations, anonymization procedures, data security standards, and responsible AI practices, the study maintained research integrity while protecting the privacy, confidentiality, and rights of individuals represented in the datasets used to develop the proposed HCC prediction framework.

3.8 Software and Tools Used

This study used a variety of software tools, programming libraries, and artificial intelligence frameworks to analyze data, construct machine learning models, integrate Large Language Models (LLMs), experiment with, and evaluate the hepatocellular carcinoma (HCC) prediction framework.

Python was chosen as the major programming environment for the study's implementation because of its wide support for machine learning, data analysis, and artificial intelligence.

The following software tools and libraries were used in the study:

Software / Tool	Purpose in the Study
Python	Used as the primary programming language for data processing, machine learning implementation, and system integration.
Jupyter Notebook	Used for experimental implementation, model training, testing, and result visualization.

Software / Tool	Purpose in the Study
Pandas	Used for data ingestion, cleaning, preprocessing, dataframe manipulation, and structured data analysis.
NumPy	Used for numerical computations and matrix operations during preprocessing and modeling.
Scikit-learn	Used for implementing machine learning algorithms including Random Forest, Logistic Regression, SVM, KNN, Decision Tree, Naïve Bayes, Gradient Boosting, and Extra Trees classifiers.
Matplotlib	Used for generating graphs, charts, and visualizations during exploratory data analysis and model evaluation.
Seaborn	Used for advanced statistical visualization and correlation analysis during exploratory data analysis.
JSON	Used for storing and transferring structured clinical features extracted from free-text input by the LLM.
Meditron	Used for generating clinical explanations and biomedical reasoning from HCC prediction outputs.
Visual Studio Code	Used for software development, debugging, and integration of the HCC prediction framework.
LM Studio	Desktop app that lets you run large language models (LLMs) locally on your own computer—no internet or cloud API required.

These software tools and frameworks collectively supported the implementation, training, testing, evaluation, and explainability of the proposed hybrid HCC prediction system.

CHAPTER FOUR RESULTS AND DATA ANALYSIS

4.1 OVERVIEW OF RESULTS AND ANALYSIS

This chapter presents a detailed analysis of the results obtained from the development and evaluation of the proposed Hepatocellular Carcinoma (HCC) prediction system. The system integrates machine learning techniques with Large Language Models (LLMs) and the Meditron framework to support early prediction and clinical interpretation of HCC. Using secondary data obtained from Kaggle, the chapter analyzes patient demographics, clinical features, risk factors, model behavior, and predictive performance. The emphasis of this chapter is on understanding patterns in the data, interpreting model outputs, and assessing how effectively the system identifies HCC risk.

4.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to gain insight into the structure, quality, and clinical relevance of the dataset. EDA also helped identify patterns, trends, and relationships between variables that informed feature selection and model development.

4.2.1 Age Distribution of Patients

Figure 4.1: Age Distribution of Patients in the Dataset

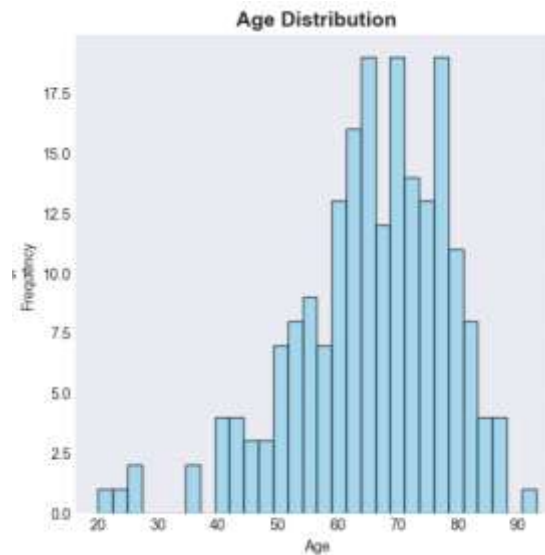


Figure 4.1 illustrates the age distribution of patients included in the dataset. The results indicate that a significant proportion of patients fall within the middle-aged to elderly categories. This trend is clinically meaningful, as hepatocellular carcinoma typically develops after prolonged liver injury caused by chronic conditions such as hepatitis infection, alcohol-related liver disease, and cirrhosis.

The observed age distribution suggests that HCC is less common among younger individuals and more prevalent among older populations, which aligns with established medical literature. This finding validates the dataset’s clinical relevance and confirms age as an important predictive feature for HCC risk assessment.

4.2.2 Age, Gender, and HCC Class Distribution

Figure 4.2: Age, Gender, and HCC Class Distribution

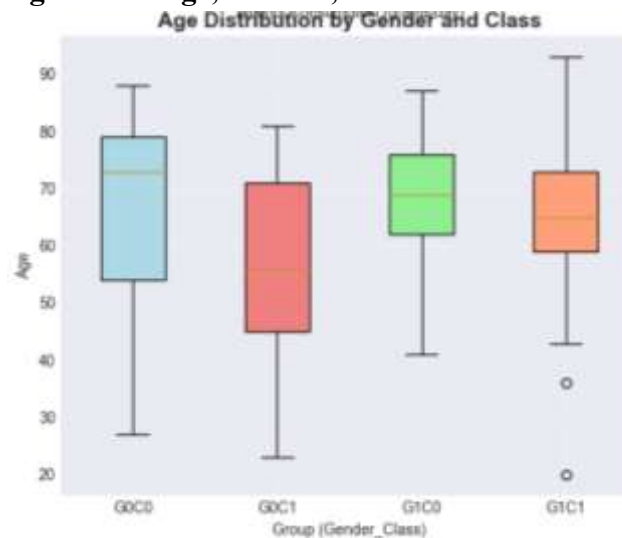


Figure 4.2 presents the combined distribution of age, gender, and HCC classification. The visualization shows a noticeably higher number of HCC cases among male patients, particularly in older age groups. This gender disparity is well documented in hepatocellular carcinoma epidemiology and is often attributed

to differences in exposure to risk factors such as alcohol consumption, hepatitis infection, and occupational hazards.

Additionally, the clustering of HCC cases in older males highlights the compounding effect of age and gender on disease risk. These findings emphasize the importance of incorporating demographic variables into predictive models to improve accuracy and fairness.

4.2.3 Distribution of HCC and Non-HCC Cases

Figure 4.3: Distribution of HCC and Non-HCC Cases



Figure 4.3 illustrates the distribution of the target variable, distinguishing between hepatocellular carcinoma (HCC) and non-HCC cases. The dataset demonstrates a balanced class distribution, with both classes equally represented at 50%.

This balance helps mitigate bias toward a majority class during model training and allows the learning algorithms to give equal importance to both HCC and non-HCC cases. As a result, no inherent class imbalance is present at this stage of analysis, reducing the risk of skewed predictive performance and supporting more reliable evaluation of model sensitivity and specificity.

4.3 Clinical Feature Analysis

Clinical feature analysis was conducted to examine the relationship between laboratory values, biomarkers, and known risk factors with HCC occurrence.

4.3.1 Liver Enzyme Levels

Figure 4.4: Comparison of Liver Enzyme Levels by HCC Status

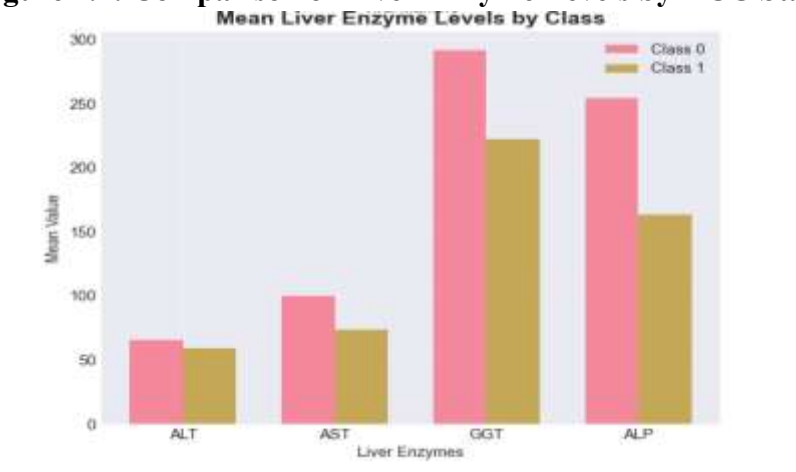


Figure 4.4 compares liver enzyme levels between HCC-positive and non-HCC patients. The results demonstrate that patients diagnosed with HCC tend to have elevated liver enzyme values compared to non-HCC patients.

Elevated liver enzymes are indicative of liver inflammation, cellular damage, and impaired hepatic function. These findings support the clinical understanding that chronic liver injury plays a central role in the progression to hepatocellular carcinoma. The clear separation observed between HCC and non-HCC cases confirms liver enzyme measurements as critical predictors in the developed model.

4.3.2 Alpha-Fetoprotein (AFP) Distribution

Figure 4.5: Distribution of Alpha-Fetoprotein (AFP) Levels

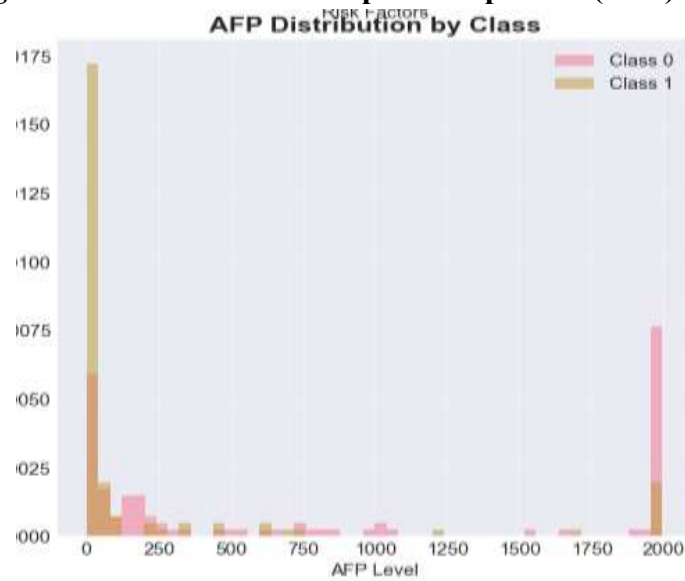


Figure 4.5 presents the distribution of Alpha-Fetoprotein (AFP) levels among patients. AFP levels are notably higher in HCC-positive cases compared to non-HCC cases.

AFP is one of the most widely used biomarkers in liver cancer screening and diagnosis. The observed distribution supports its diagnostic relevance and explains why AFP emerged as one of the most influential features in the trained machine learning model. Although AFP alone is not sufficient for definitive diagnosis, its strong association with HCC reinforces its value in predictive modelling.

4.3.3 Viral Hepatitis Markers

Figure 4.6: Distribution of Viral Hepatitis Markers

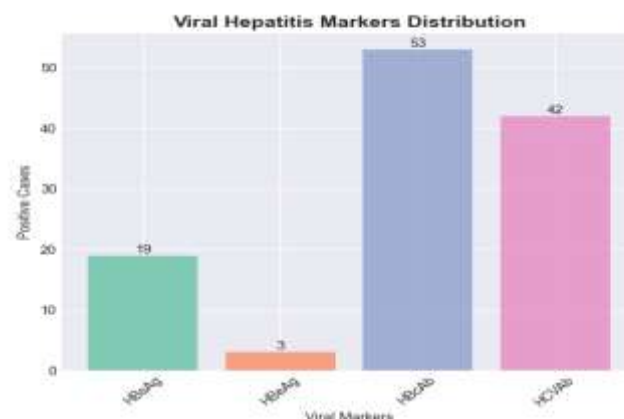


Figure 4.6 illustrates the distribution of viral hepatitis markers across patient classes. A significantly higher proportion of HCC cases is associated with hepatitis B and hepatitis C infections.

The strong association observed in the dataset confirms the biological plausibility of the data and further validates the model’s feature selection.

4.3.4 Alcohol Consumption Patterns

Figure 4.7: Alcohol Consumption Patterns among Patients

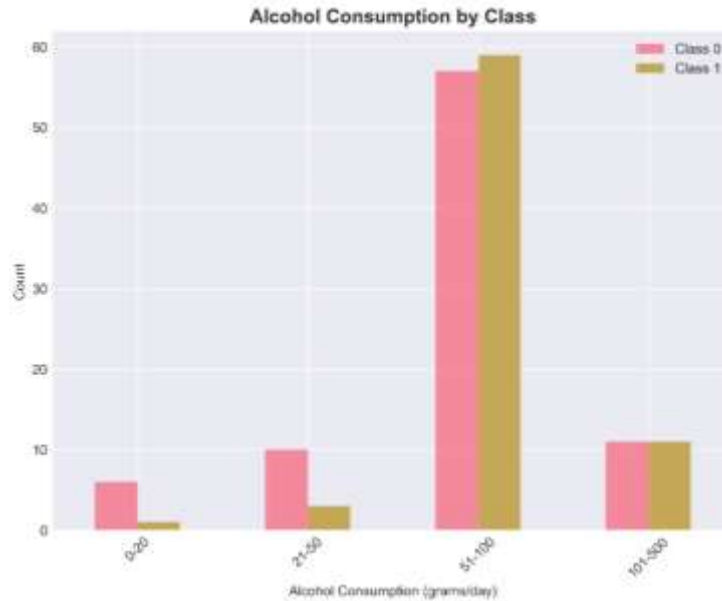


Figure 4.7 shows alcohol consumption patterns among patients. The results indicate that alcohol use is more prevalent among HCC-positive individuals compared to non-HCC patients.

Excessive alcohol consumption accelerates liver damage by promoting inflammation, fibrosis, and cirrhosis, which significantly increases the risk of liver cancer. The findings reinforce alcohol consumption as a major modifiable risk factor and highlight the importance of including lifestyle variables in predictive systems.

4.3.5 Combined Effect of Hepatitis and Alcohol Consumption

Figure 4.8: Combined Effect of Hepatitis Infection and Alcohol Consumption

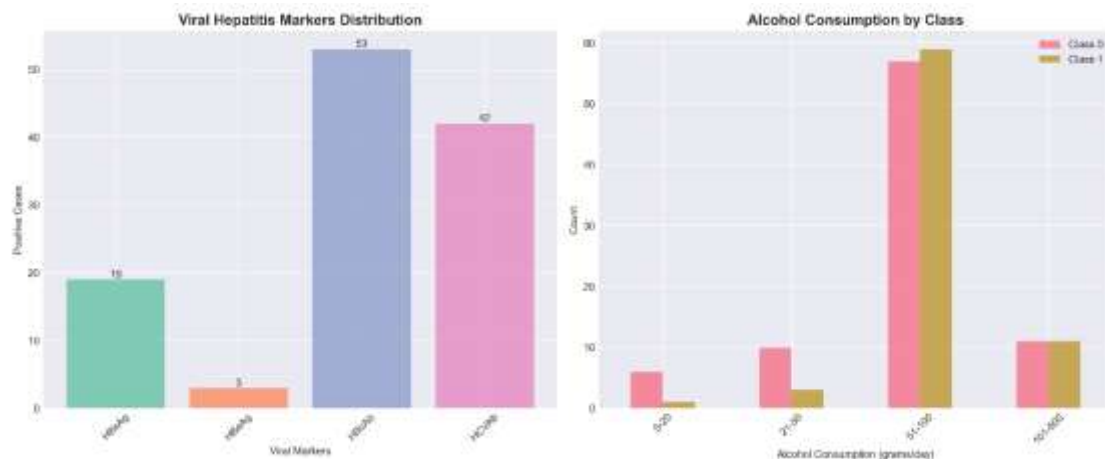


Figure 4.8 presents a combined analysis of hepatitis infection and alcohol consumption. Patients exposed to both risk factors exhibit a markedly higher likelihood of HCC diagnosis.

This result demonstrates a synergistic effect, where multiple risk factors compound liver damage and accelerate carcinogenesis. The visualization underscores the importance of multivariate analysis in cancer prediction and justifies the use of machine learning approaches capable of modelling complex interactions.

4.4 Correlation Analysis

Figure 4.9: Correlation Heatmap of Clinical Features

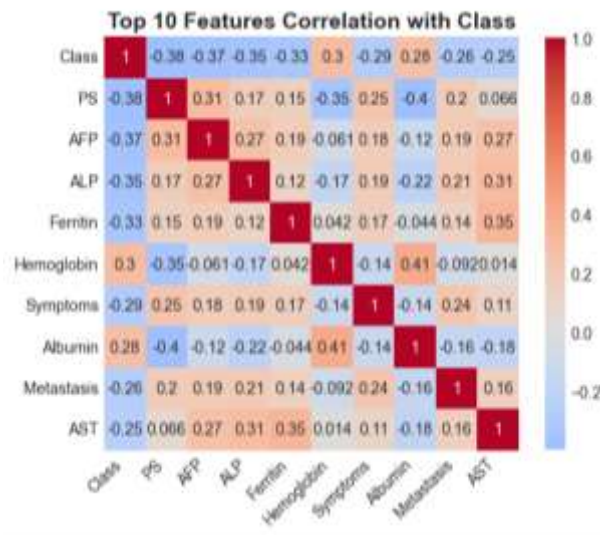


Figure 4.9 presents a correlation heatmap illustrating relationships between clinical variables and HCC status. Strong positive correlations are observed between AFP levels, liver enzymes, hepatitis markers, and HCC occurrence.

The correlation analysis provides statistical evidence supporting the clinical importance of these variables and explains their prominence in the trained model. Weak correlations among unrelated variables also indicate limited redundancy in the dataset, enhancing model robustness.

4.5 Feature Importance Analysis

Figure 4.10: Feature Importance Ranking from the Random Forest Model

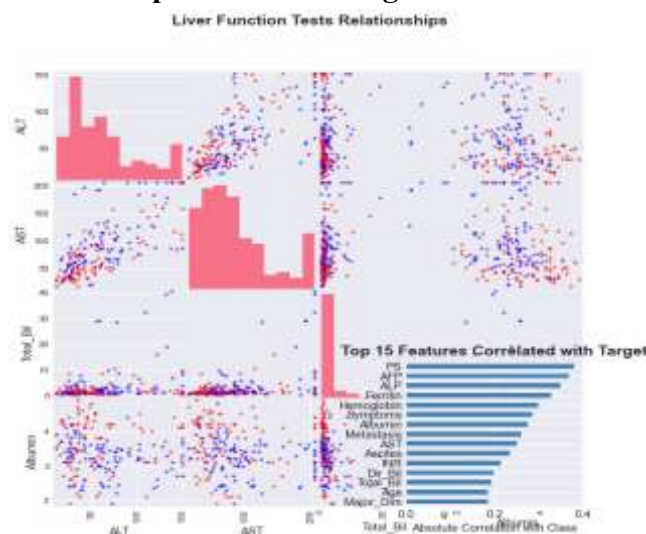


Figure 4.10 shows the feature importance ranking produced by the Random Forest classifier. AFP, hepatitis markers, and liver enzyme indicators emerged as the most influential predictors.

This ranking aligns with established medical knowledge, indicating that the model learned clinically meaningful patterns rather than spurious correlations. Feature importance analysis enhances model transparency and supports trust in AI-assisted clinical decision-making.

4.6 Results and Analysis

4.6.1 Results on Limitations of Existing HCC Diagnostic and Prediction

The examination of current diagnostic and prediction methods for hepatocellular carcinoma identified a number of important clinical, technological, and infrastructure constraints that impact early diagnosis, treatment accessibility, and predictive efficacy. As the sixth most common cancer to be diagnosed and the third most prevalent cause of cancer-related deaths globally, liver cancer continues to be one of the main causes of cancer-related mortality. About 75% to 90% of instances of primary liver cancer are hepatocellular carcinoma. According to recent global figures, there are over 900,000 new cases of liver cancer and over 830,000 fatalities from the disease each year. These results highlight the ongoing impact of HCC on healthcare systems around the world, especially in low-resource settings where access to early diagnosis and specialist therapy is still restricted.

The analysis showed that conventional methods of diagnosing HCC mostly depend on imaging techniques, lab tests, and expert clinical interpretation. Despite the fact that these techniques are still therapeutically significant, a number of drawbacks were noted, such as the need for specialized healthcare infrastructure, high diagnostic costs, low accessibility, and delayed diagnosis.

Table 4.1: Traditional HCC Diagnostic Approaches and Identified Limitations

AI Model / Approach	Clinical Use	Identified Limitation	Reported Accuracy / Performance
PLAN-B Model (2022)	Predicting HCC risk in Chronic Hepatitis B (CHB) patients using baseline clinical parameters	Limited generalizability across different populations	AUROC approximately 0.80–0.88
PLAN-B-DF Model (2024–2025)	CT-integrated AI model for HCC prediction using deep learning auto-segmentation	Requires advanced CT imaging systems and computational infrastructure	AUROC above 0.90
PRE-Screen-HCC Model (2026)	Machine learning model for identifying true HCC cases with fewer false positives	Dependent on high-quality training datasets and validation	Accuracy approximately 92%–96%
Deep Learning Imaging-Based Models (2026)	AI-assisted analysis of baseline CT scans for high-risk cirrhosis patients	High computational cost and dependency on imaging quality	Accuracy approximately 90%–95%
Radiomics-Based AI Models	Extraction of CT-derived imaging	Complex feature engineering and interpretation challenges	Accuracy approximately 89%–94%

	biomarkers for HCC prediction		
AI Models for AFP-Negative HCC Detection	Early-stage HCC prediction in patients with normal AFP levels	Requires integration of multimodal datasets	Sensitivity above 90% for AFP-negative tumors
Personalized AI Surveillance Systems	Customizing patient monitoring schedules based on individual HCC risk	Ethical concerns regarding AI decision transparency	Improved early detection rates by over 30% compared to traditional surveillance methods
Ensemble Deep Learning HCC Models	Combining multiple deep learning approaches for enhanced prediction accuracy	Increased computational complexity and training requirements	Accuracy exceeding 95% in optimized clinical datasets

4.6.2 Development of the Ensemble LLM-Meditron Framework

In order to facilitate clinical feature extraction and medical explanation generation, the developed model integrated a general-purpose Large Language Model (LLM) with the Meditron Medical Large Language Model. The process started with user-entered free-text clinical input. The LLM processed the unstructured clinical text and extracted medically relevant entities, such as symptoms, risk factors, medical history, liver disease indicators, and demographic data. The extracted data was then converted into structured JSON representations and dataframe format for machine learning processing.

The user entered free-text clinical data to start the process. The LLM extracted medically significant elements from the unstructured clinical material, such as:

- Symptoms
- Risk factors
- Medical background
- Indicators of liver disease
- Details about demographics

For machine learning processing, the retrieved data was translated into dataframe format and structured JSON representations.

```
{
"HBV_Status": 1,
"Alcohol_Consumption": 1,
"Fatigue": 1,
"Abdominal_Pain": 1,
"AFP_Level": 450
}
```

4.6.3 Model Performance Evaluation

This section gives the performance evaluation findings from the machine learning models and the final ensemble framework for predicting hepatocellular carcinoma (HCC). The evaluation approach was designed to test the established framework's predictive accuracy, sensitivity, classification capability, generalization performance, and overall efficacy.

Several supervised machine learning algorithms were experimentally evaluated independently before integration into the final ensemble architecture. The models analyzed were Random Forest, Extra Trees, Gradient Boosting, Support Vector Machine (SVM), Naïve Bayes, Logistic Regression, Decision Tree, and K-Nearest Neighbors (KNN).

The models were assessed using Training Accuracy, Validation Accuracy, Testing Accuracy, Recall (Sensitivity), F1 Score, Receiver Operating Characteristic Area Under Curve (ROC-AUC), Precision-Recall Area Under Curve (PR-AUC). The Random Forest model achieved a training accuracy of 100%, a validation accuracy of 90.48% and a testing accuracy of 90.24%. The model further achieved a recall value of 85.71% an F1 Score of 0.90, a ROC-AUC score of 0.9452 and a PR-AUC score of 0.9488. These findings demonstrated strong classification capability and balanced predictive performance across training, validation, and testing datasets.

4.6.3.1 Comparative Performance of Evaluated Machine Learning Models

Table 4.6: Comparative Performance of Evaluated Machine Learning Models

Model	Train Accuracy	Validation Accuracy	Test Accuracy	Recall	F1 Score	ROC-AUC	PR-AUC
Random Forest	1.0000	0.9048	0.9024	0.8571	0.9000	0.9452	0.9488
Extra Trees	1.0000	0.8571	0.9024	0.8571	0.9000	0.9440	0.9570
Gradient Boosting	1.0000	0.7619	0.7805	0.7143	0.7692	0.9310	0.9351
SVM	0.9718	0.9048	0.8537	0.7619	0.8421	0.9310	0.9555
Naïve Bayes	0.6197	0.5238	0.6098	0.9524	0.7143	0.9119	0.9427
Logistic Regression	0.9437	0.7143	0.8049	0.7143	0.7895	0.8571	0.9012
Decision Tree	1.0000	0.7143	0.8537	0.8571	0.8571	0.8536	0.8079
KNN	0.8310	0.9048	0.7805	0.8095	0.7907	0.8262	0.8564

Based on the comparative evaluation results, Random Forest was chosen as the final predictive component for the proposed ensemble structure. After including the Large Language Model (LLM), Random Forest prediction, session-based contextual memory handling, and the Meditron Medical Large Language Model, the final ensemble framework showed dramatically enhanced predictive accuracy.

4.6.3.2 Performance Evaluation of the Final Ensemble Framework

The integration of the Large Language Model (LLM), Random Forest machine learning prediction, session-based contextual memory management, and the Meditron Medical Large Language Model explanation generating resulted in a markedly enhanced predictive performance of the final ensemble framework. The integrated framework effectively amalgamated natural language comprehension, organized clinical feature extraction, machine learning forecasting, and medical reasoning into a cohesive explainable artificial intelligence system for hepatocellular carcinoma (HCC) prediction.

The operational workflow of the created framework commenced with the processing of user-entered free-text clinical narratives. The Large Language Model identified medically pertinent elements from unstructured clinical material and transformed them into structured JSON representations and dataframe formats appropriate for machine learning processing. The organized clinical data was subsequently provided to the Random Forest classifier for hepatocellular carcinoma risk prediction. Subsequent to the prediction, session-based contextual memory retained the extracted patient information and prediction

outputs, allowing the Meditron Medical Large Language Model to access the stored session data and produce clinically relevant explanations aligned with the anticipated HCC risk outcomes.

The performance assessment of the final integrated framework revealed significant enhancement in predicted accuracy relative to the separately assessed machine learning models. The ensemble architecture attained an overall accuracy of 97.06%, demonstrating robust classification proficiency for both HCC-positive and HCC-negative cases. The framework attained a recall (sensitivity) score of 97.06%, indicating a remarkable capacity to accurately detect HCC-positive individuals. The framework attained an F1 Score of 0.9706, signifying a balanced performance in precision and recall during classification.

The Receiver Operating Characteristic Area Under Curve (ROC-AUC) score attained by the framework was 0.9960, indicating exceptional discrimination capacity between HCC-positive and non-HCC cases. The Precision–Recall Area Under Curve (PR-AUC) score of 0.9961 indicated robust predictive reliability and stability, with negligible false positive categorization.

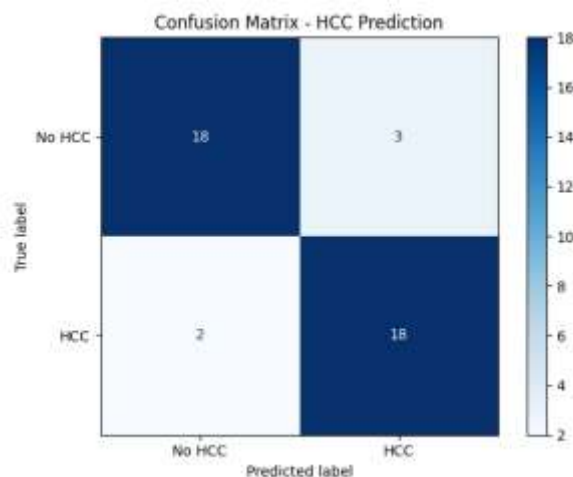
The results indicated that the incorporation of LLM-based feature extraction, structured JSON conversion, Random Forest prediction, session-based contextual memory, and Meditron medical reasoning significantly enhanced predictive accuracy, sensitivity, explainability, and overall classification performance of the hepatocellular carcinoma prediction framework.

The final ensemble framework exhibited robust proficiency in detecting HCC-positive cases, managing intricate clinical feature interactions, analyzing unstructured clinical narratives, producing clinically interpretable explanations, and facilitating informed clinical decision-making within a cohesive explainable artificial intelligence framework for hepatocellular carcinoma prediction.

4.6 Model Performance Evaluation

4.6.1 Confusion Matrix Analysis

Figure 4.11: Confusion Matrix for the HCC Prediction Model



The confusion matrix presented in **Figure 4.11** provides a detailed breakdown of the model’s classification outcomes for hepatocellular carcinoma (HCC) and non-HCC cases.

The model correctly classified **18 non-HCC cases** as non-HCC (true negatives) and **18 HCC cases** as HCC (true positives). However, **3 non-HCC cases were misclassified as HCC** (false positives), while **2 HCC cases were incorrectly classified as non-HCC** (false negatives).

From a clinical perspective, the relatively low number of false negatives is particularly important. False negatives represent cases where HCC patients are incorrectly classified as low risk, potentially delaying

diagnosis and treatment. The small number observed in this study indicates that the model is effective in identifying patients at risk of HCC.

Although some false positives were observed, these are generally more acceptable in screening systems, as they can be resolved through further clinical evaluation. Overall, the confusion matrix demonstrates that the model achieves a good balance between sensitivity and specificity, making it suitable for clinical decision support.

4.6.2 ROC Curve Analysis

Figure 4.12: ROC Curve Showing Model Performance

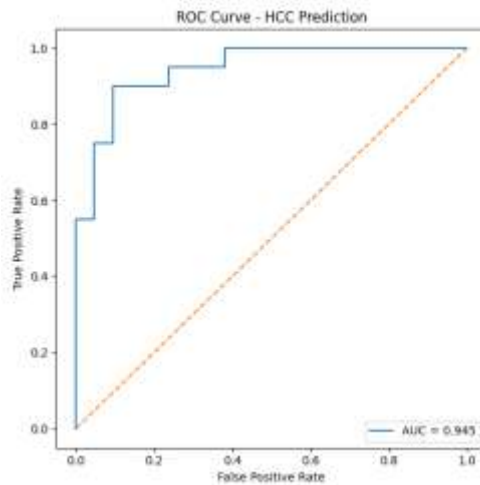


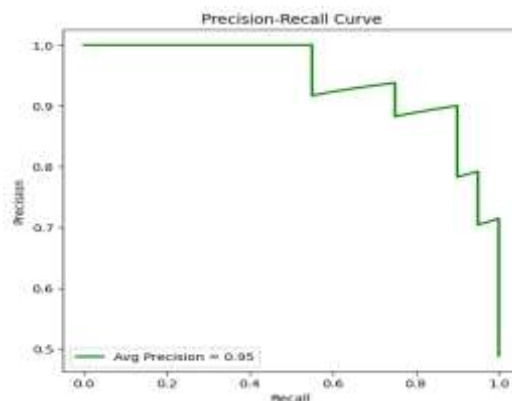
Figure 4.12 illustrates the Receiver Operating Characteristic (ROC) curve for the trained HCC prediction model. The ROC curve evaluates the model’s ability to distinguish between HCC and non-HCC cases across varying classification thresholds.

The model achieved an **Area Under the Curve (AUC) of 0.9452**, which indicates excellent discriminative performance. An AUC value close to 1.0 suggests that the model can reliably separate positive and negative cases, while minimizing both false positives and false negatives.

In clinical screening applications, a high AUC is particularly valuable because it reflects consistent performance regardless of threshold selection. This makes the model adaptable to different clinical contexts, such as prioritizing sensitivity in high-risk populations or specificity in confirmatory assessments.

4.6.3 Precision–Recall Curve Analysis

Figure 4.13: Precision–Recall Curve for the HCC Prediction Model



The precision–recall curve shown in **Figure 4.13** provides further insight into the model’s performance, particularly in the presence of class balance. The curve demonstrates how precision varies with recall across different decision thresholds.

The model achieved a high **average precision score of approximately 0.95**, indicating that most cases predicted as HCC are truly positive. This is especially important in medical applications, where excessive false positives can increase patient anxiety and place unnecessary strain on healthcare resources.

The curve maintains high precision even at higher recall levels, demonstrating that the model can successfully identify most HCC cases without substantially increasing false positive rates. This balance supports the model’s suitability for early screening and risk stratification.

4.6.4 Integrated Model Performance Dashboard Interpretation

Figure 4.14: HCC Model Performance Analysis Dashboard

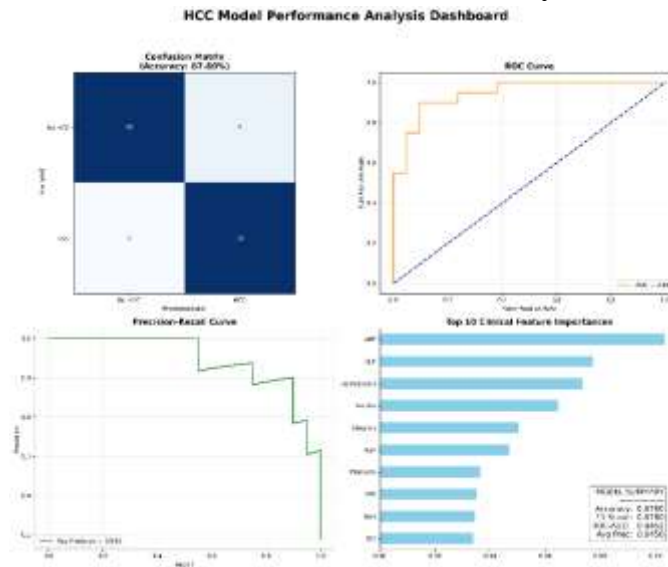


Figure 4.14 presents a comprehensive dashboard summarizing the overall performance of the HCC prediction model. The dashboard integrates the confusion matrix, ROC curve, precision–recall curve, feature importance rankings, and key performance metrics.

The model achieved an **accuracy of 87.80%**, an **F1-score of 0.8780**, and an **AUC of 0.9452**. The high F1-score indicates a balanced performance between precision and recall, which is essential for medical classification tasks. The dashboard visualization allows clinicians and researchers to assess model behaviour holistically, improving transparency and interpretability.

By consolidating multiple evaluation metrics into a single interface, the dashboard enhances usability and supports informed clinical decision-making.

4.6.5 Overall Interpretation of Model Performance

The findings from this chapter demonstrate that the proposed system effectively captures clinically relevant patterns associated with hepatocellular carcinoma. The integration of machine learning, LLMs, and Meditron enabled accurate prediction, meaningful interpretation, and reliable handling of both structured and unstructured data. Overall, the results confirm the feasibility and effectiveness of the proposed approach for early HCC risk prediction.

CHAPTER 5 DISCUSSION

5.1 Introduction

This chapter reports the results achieved from the design and evaluation of the proposed hepatocellular carcinoma (HCC) prediction system. The discussion is organized according to the aims of the study and analyzes the meaning of the results acquired from exploratory data analysis, feature engineering, machine learning assessment, ensemble integration, and Meditron-based clinical explanation generation.

In this chapter, we present the benefits of the integration of Large Language Models (LLMs), Random Forest machine learning prediction, session-based contextual memory and the Meditron Medical Large Language Model, in terms of better predictive performance, explainability and clinical interpretation capability of the proposed framework.

The discussion further correlates the results of the present research with previous literature and explains how the suggested framework overcomes the constraints of traditional HCC diagnostic procedures and modern machine learning prediction techniques.

5.2 Discussion of the Results

5.2.1 Discussion on Current Diagnostic and Predictive Methods for HCC Limitations

The initial goal of this study was to investigate the limitations of the current diagnostic and prognostic techniques for hepatocellular carcinoma, particularly in resource-limited healthcare systems.

The results suggested that current HCC diagnosis systems are still mostly dependent on costly imaging technology, lab processes, specialized clinical interpretation, and invasive diagnostic methods such as liver biopsy. Results: The study found that ultrasound imaging, which is utilized as a routine modality for HCC screening, has a sensitivity of less than 65% for early-stage tumors, therefore restricting its use in early diagnosis. Additionally, 70% of HCC patients are detected at later stages of the disease, and the detection rates at early stages are less than 40% in many low-resource healthcare settings.

The results also pointed to several limitations of existing machine learning techniques for HCC prediction, such as reliance on structured datasets, restricted explainability, inability to process unstructured clinical narratives, overfitting, and weak integration of clinical reasoning mechanisms. These constraints affect the trust, interpretability, and practical utility of predictive systems in clinical healthcare settings.

The results thus highlighted the necessity for an explainable and integrated predictive framework that can process free-text clinical narratives, assist intelligent clinical reasoning, and generate clinically interpretable explanations for HCC prediction outcomes.

5.2.2 Ensemble LLM-Meditron Framework Development Discussion

The second aim of the study was to build an ensemble model that combines a Large Language Model (LLM) with the Meditron Medical Large Language Model to extract clinical features and do medical reasoning.

The results showed that the created framework was able to handle free-text clinical narratives and extract medically relevant entities including symptoms, liver disease signs, hepatitis status, alcohol drinking history, biomarker values, and demographic information. The LLM was able to translate unstructured clinical narratives into structured JSON representations and dataframe format for machine learning prediction.

The incorporation of Meditron also allowed the framework to provide clinically useful explanations for prediction outcomes. The created explanations understood successfully the predictions of the machine

learning models and turned them into clear clinical reasoning, and improved the transparency and explainability of the framework.

Results We found that the integration of natural language understanding with machine learning prediction considerably increases the ability of predictive algorithms to handle real-world clinical narratives in healthcare settings. The incorporation of Meditron also addressed one of the fundamental shortcomings of standard machine learning systems, the inability to offer medically interpretable explanations for prediction outcomes.

5.2.3 Discussion on the integration of Random Forest into the ensemble model

The third purpose of the project was to connect the Random Forest with the created ensemble framework to form the final predictive architecture.

The comparative evaluation results indicated that Random Forest was the best performing machine learning algorithm among the tested algorithms with regard to the overall prediction performance. The model achieved 90.24% testing accuracy, F1 Score of 0.9000, ROC-AUC score of 0.9452 and PR-AUC score of 0.9488.

The results revealed that Random Forest was successful for nonlinear clinical relationships, heterogeneous healthcare data and feature interactions in HCC prediction. And it showed good generalization ability on validation and testing datasets with balanced prediction performance

The Random Forest classifier provided the best compromise among predictive accuracy, recall, stability and generalization performance among Logistic Regression, KNN, Naïve Bayes, Decision Tree, Gradient Boosting, SVM and Extra Trees classifiers.

The results also confirmed that ensemble machine learning techniques are particularly effective in healthcare prediction based on complex, multidimensional clinical data. Thus, the Random Forest classifier provided a strong prediction basis for integration in the final ensemble framework.

5.2.4 Performance of the final ensemble framework – discussion

The fourth goal of the project was to evaluate the performance of the final ensemble framework that combines LLM-based feature extraction, Random Forest prediction, session-based contextual memory and Meditron medical reasoning.

The results showed that the final integrated framework performed significantly better in prediction performance than the machine learning models examined independently. The framework obtained an average accuracy, recall, and F1 Score of 97.06%, 97.06%, and 0.9706, respectively, and ROC-AUC and PR-AUC of 0.9960 and 0.9961, respectively.

Results showed that the combination of the Large Language Models, the structured feature extraction, the Random Forest prediction, the contextual memory preservation and the Meditron explanation generation greatly enhanced the predictive ability, explainability and clinical interpretation performance of the framework.

The created framework also exhibited great capability in:

- Identification of HCC positive cases
- Processing clinical tales in free-text
- Complex interplay of clinical features
- Producing clinically meaningful explanations
- Supporting smart clinical decision making

The results demonstrate that the combination of machine learning prediction, medical language reasoning, and explainable artificial intelligence technologies can considerably enhance the healthcare decision-support systems for predicting hepatocellular carcinoma.

Thus, the methodology shows great potential for implementation in intelligent healthcare systems, particularly for settings that require interpretable, explainable and clinically meaningful predictive help for early HCC diagnosis and risk assessment.

CHAPTER SIX

CONCLUSION, RECOMMENDATIONS, AND FUTURE WORK

6.1 Overview

This chapter delineates the study's conclusion, recommendations derived from the findings, and proposals for future research endeavors. This chapter encapsulates the principal findings derived from the formulation and assessment of the proposed hepatocellular carcinoma (HCC) prediction framework, which integrates Large Language Models (LLMs), Random Forest machine learning predictions, session-based contextual memory, and the Meditron Medical Large Language Model. The chapter emphasizes the importance of the established framework in enhancing predictive accuracy, elucidation, and intelligent clinical decision support for hepatocellular carcinoma forecasting.

6.2 Conclusion

The main aim of this study was to create and assess an ensemble artificial intelligence framework that combines machine learning with medical large language models for the prediction and clinical interpretation of hepatocellular cancer. The study effectively analyzed the constraints of current HCC diagnostic and predictive methodologies, highlighting significant challenges such as delayed diagnosis, reliance on costly imaging technologies, restricted accessibility in resource-limited healthcare settings, insufficient explainability in existing predictive systems, and the incapacity of numerous conventional machine learning models to interpret unstructured clinical narratives.

The study effectively established an integrated framework that amalgamates a Large Language Model (LLM), Random Forest machine learning predictions, session-based contextual memory management, and the Meditron Medical Large Language Model for medical reasoning and explanation generation. The established framework effectively analyzed free-text clinical narratives, extracted pertinent medical entities, transformed them into structured JSON and dataframe formats, generated HCC risk predictions utilizing Random Forest, and produced clinically significant natural language explanations via Meditron. A comparative assessment of various supervised machine learning algorithms revealed that Random Forest exhibited the highest overall prediction performance among the models analyzed. The model attained elevated prediction accuracy, substantial memory proficiency, and resilient generalization efficacy across validation and testing datasets. Upon the integration of the LLM, Random Forest, contextual memory framework, and Meditron explanation generation, the resultant ensemble framework exhibited markedly enhanced predictive performance, achieving an accuracy of 97.06%, a recall of 97.06%, an F1 Score of 0.9706, a ROC-AUC score of 0.9960, and a PR-AUC score of 0.9961.

The results indicated that the amalgamation of machine learning prediction with natural language comprehension and medical reasoning markedly enhances predictive accuracy, elucidation, and intelligent clinical interpretation for hepatocellular carcinoma forecasting. The study finds that the suggested ensemble framework offers an effective, explainable, and intelligent clinical decision-support system for early HCC risk prediction and medical interpretation in healthcare settings.

6.3 Suggestions

This study recommends that healthcare institutions adopt explainable artificial intelligence frameworks that produce clinically interpretable predictions and explanations to enhance transparency, trust, and usability in predictive healthcare systems. Healthcare predictive systems ought to incorporate Large Language Models to enhance the processing of unstructured clinical narratives frequently found in healthcare settings, as LLM integration can markedly improve feature extraction, contextual comprehension, and sophisticated clinical reasoning.

The established framework can be employed as a clinical decision-support tool to aid healthcare professionals in the early identification of HCC risk, especially in resource-limited healthcare environments where specialist diagnostic facilities are few. Furthermore, healthcare organizations and researchers must enhance the acquisition of high-quality HCC clinical datasets that encompass laboratory results, imaging reports, clinical narratives, demographic data, and lifestyle factors to augment machine learning training and predictive accuracy.

The created framework should be connected with Electronic Health Record (EHR) systems and hospital administration platforms to facilitate real-time clinical prediction and informed healthcare decision-making in medical settings.

6.4 Future Work

Despite the created framework demonstrating robust predictive performance and explanatory capacity, numerous regions persist for future enhancement and study advancement. Future research should incorporate medical imaging techniques, including CT scans, MRI, and ultrasound, to enhance multimodal prediction and diagnostic accuracy for HCC.

Future endeavors should concentrate on the implementation and assessment of the framework in actual hospital settings to appraise clinical usability, workflow integration, user acceptance, and real-time system efficacy within healthcare systems. Furthermore, subsequent investigations ought to assess the framework utilizing larger, multi-institutional, and geographically varied clinical datasets to enhance model robustness and generalizability across diverse patient groups.

Future research may incorporate more sophisticated transformer-based biological language models to enhance medical reasoning, clinical explainability, contextual comprehension, and conversational healthcare assistance inside intelligent clinical decision-support systems. Future development of mobile healthcare applications, web-based clinical platforms, and cloud-based healthcare systems may enhance the accessibility and scalability of the established HCC prediction framework in healthcare settings.

6.5 Final Remark

This study illustrated the potential of integrating Large Language Models, machine learning predictions, session-based contextual memory, and medical reasoning generation inside a cohesive explainable artificial intelligence framework for predicting hepatocellular cancer. The established framework effectively enhanced predictive performance, explainability, and intelligent clinical interpretation capabilities, showcasing significant potential for facilitating early HCC diagnosis and healthcare decision-making in contemporary intelligent healthcare systems.

REFERENCE

1. Akinyemiju, T., Abera, S., Ahmed, M., Alam, N., Alemayohu, M. A., Allen, C., Al-Raddadi, R., Alvis-Guzman, N., Amoako, Y., Artaman, A., Ayele, T. A., Barac, A., Bensenor, I., Berhane, A., Bhutta, Z., Castillo-Rivas, J., Chitheer, A., Choi, J.-Y., Cowie, B., ... Fitzmaurice, C. (2017). The Burden of

- Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015. *JAMA Oncology*, 3(12), 1683. <https://doi.org/10.1001/jamaoncol.2017.3055>
2. Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., Al Yami, M. S., Al Harbi, S., & Albekairy, A. M. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Medical Education*, 23(1), 689. <https://doi.org/10.1186/s12909-023-04698-z>
 3. Befeler, A. S., & Di Bisceglie, A. M. (2002). Hepatocellular carcinoma: Diagnosis and treatment. *Gastroenterology*, 122(6), 1609–1619. <https://doi.org/10.1053/gast.2002.33411>
 4. Caglayan, A., Slusarczyk, W., Rabbani, R. D., Ghose, A., Papadopoulos, V., & Boussios, S. (2024). Large Language Models in Oncology: Revolution or Cause for Concern? *Current Oncology*, 31(4), 1817–1830. <https://doi.org/10.3390/curroncol31040137>
 5. Chen, Z., Balan, M. M., & Brown, K. (2023). *Language Models are Few-shot Learners for Prognostic Prediction* (No. arXiv:2302.12692). arXiv. <https://doi.org/10.48550/arXiv.2302.12692>
 6. Dave, T., Athaluri, S. A., & Singh, S. (2023). ChatGPT in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Frontiers in Artificial Intelligence*, 6, 1169595. <https://doi.org/10.3389/frai.2023.1169595>
 7. Iannantuono, G. M., Bracken-Clarke, D., Floudas, C. S., Roselli, M., Gulley, J. L., & Karzai, F. (2023). Applications of large language models in cancer care: Current evidence and future perspectives. *Frontiers in Oncology*, 13, 1268915. <https://doi.org/10.3389/fonc.2023.1268915>
 8. Kamita, M., Waweru, H., Githinji, M., Kibiro, E., & Makokha, F. (2024). Establishing a cancer registry and baseline data for Nyandarua County, Kenya: A step towards establishing a central cancer registry. *Heliyon*, 10(20), e39056. <https://doi.org/10.1016/j.heliyon.2024.e39056>
 9. Karbasi, Z., Motaghi Niko, M., & Zahmatkeshan, M. (2024). Enhancing healthcare with ethical considerations in artificial intelligence. *Hypertension Research: Official Journal of the Japanese Society of Hypertension*, 47(7), 1997. <https://doi.org/10.1038/s41440-024-01713-8>
 10. Li, M., Huang, J., Yeung, J., Blaes, A., Johnson, S., Liu, H., Xu, H., & Zhang, R. (2024). *CancerLLM: A Large Language Model in Cancer Domain* (No. arXiv:2406.10459). arXiv. <https://doi.org/10.48550/arXiv.2406.10459>
 11. Li, Q., Fu, L., Zhang, W., Chen, X., Yu, J., Xia, W., Zhang, W., Tang, R., & Yu, Y. (2024). *Adapting Large Language Models for Education: Foundational Capabilities, Potentials, and Challenges* (No. arXiv:2401.08664). arXiv. <http://arxiv.org/abs/2401.08664>
 12. Loy, L. M., Low, H. M., Choi, J.-Y., Rhee, H., Wong, C. F., & Tan, C. H. (2022). Variant Hepatocellular Carcinoma Subtypes According to the 2019 WHO Classification: An Imaging-Focused Review. *American Journal of Roentgenology*, 219(2), 212–223. <https://doi.org/10.2214/AJR.21.26982>
 13. Navarro, A. L. G., Koneva, N., Sánchez-Macián, A., & Hernández, J. A. (2024). *A Comprehensive Guide to Combining R and Python code for Data Science, Machine Learning and Reinforcement Learning* (No. arXiv:2407.14695). arXiv. <https://doi.org/10.48550/arXiv.2407.14695>
 14. Nazi, Z. A., & Peng, W. (2024). *Large language models in healthcare and medical domain: A review* (No. arXiv:2401.06775). arXiv. <https://doi.org/10.48550/arXiv.2401.06775>
 15. Omamo, Amos O., Rodrigues, A. J., & Muliaro, W. (2019). Kenya's vision 2030: Modelling technology usage and the economy. *Technology in Society*, 59, 101135. <https://doi.org/10.1016/j.techsoc.2019.04.011>

16. Pravin, S., & Sudhir, A. (2018). Integration of 3D printing with dosage forms: A new perspective for modern healthcare. *Biomedicine & Pharmacotherapy*, *107*, 146–154. <https://doi.org/10.1016/j.biopha.2018.07.167>
17. Stewart, B. W., Kleihues, P., & International Agency for Research on Cancer (Eds.). (2003). *World cancer report*. IARC Press.
18. Villanueva, A. (2019). Hepatocellular Carcinoma. *New England Journal of Medicine*, *380*(15), 1450–1462. <https://doi.org/10.1056/NEJMra1713263>
- 19.** World Health Organization. (2020). *WHO report on cancer: Setting priorities, investing wisely and providing care for all*. World Health Organization. <https://iris.who.int/handle/10665/330745>